

Reg No.: \_\_\_\_\_

Name: \_\_\_\_\_

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**  
**EIGHTH SEMESTER B.TECH DEGREE EXAMINATION(S), OCTOBER 2019**

**Course Code: CS466**  
**Course Name: DATA SCIENCE**

Max. Marks: 100

Duration: 3 Hours

**PART A**

*Answer all questions, each carries 4 marks.*

Marks

- |    |   |     |
|----|---|-----|
| 1  | Discuss types of NoSQL databases.   | (4) |
| 2  | Write a short note on overfitting.  | (4) |
| 3  | Describe different class of problems that are best suited to be solved with machine learning. | (4) |
| 4  | Create 2 vectors x= (3,6,8) and y= (2,9,0) and add them using R. Also Compute x+1.            | (4) |
| 5  | Differentiate the functions summary() and finenum().  | (4) |
| 6  | Why area plots is important? Explain how to create an area plot in Python                     | (4) |
| 7  | List and explain any four application areas of Hadoop.  | (4) |
| 8  | How to perform matrix-vector multiplication using MapReduce?                                  | (4) |
| 9  | Outline any four deployment methods in R.   | (4) |
| 10 | How will you present your work to project sponsor?  | (4) |

**PART B**

*Answer any two full questions, each carries 9 marks.*

- |    |   |     |
|----|---|-----|
| 11 | a) Explain various methods involved in data cleaning with an example. | (9) |
| 12 | a) List William Cleveland's principles for scientific visualization   | (3) |
|    | b) How histograms and hexbin plots are created using ggplot in R      | (6) |
| 13 | a) What is k-means clustering?  | (3) |
|    | b) Write R code for implementing k-means clustering?                  | (6) |

**PART C**

*Answer any two full questions, each carries 9 marks.*

- |    |  |     |
|----|--|-----|
| 14 | a) Illustrate read.table() in R  | (4) |
|    | b) Compare lists and arrays in R? Write R codes for constructing, modifying and concatenating lists with examples. | (5) |
| 15 | a) With the help of a case study, Explain item based collaborative filtering in Python.                            | (9) |
| 16 | a) Discuss simple linear regression. Implement it by using python.   | (5) |

- b) Why trellis plot is important? Explain how to create a trellis plot in Python (4)

**PART D**

*Answer any two full questions, each carries 12 marks.*

- 17 a) What are the components of Hadoop? (8)  
b) How does master slave architecture work in the Hadoop? (4)
- 18 a) Write a MapReduce program for word count implementation via Hadoop framework? (8)  
b) Explain the usage of plot() function with an example. (4)
- 19 a) How do you export graphs in R? (6)  
b) How to do multiple plots in one window (6)

\*\*\*\*